Differential DNA Methylation in Umbilical Cord Blood of Infants Exposed to Low Levels of Arsenic *in Utero*

Devin C. Koestler, Michele Avissar-Whiting,
E. Andres Houseman, Margaret R. Karagas
and Carmen J. Marsit

Differential DNA Methylation in Umbilical Cord Blood of Infants Exposed to Low Levels of Arsenic *in Utero*

Devin C. Koestler[1], Michele Avissar-Whiting[2], E. Andres Houseman[3], Margaret R. Karagas[1], and Carmen J. Marsit[1,4]

[1] Department of Community and Family Medicine, Geisel School of Medicine at   Dartmouth College, Lebanon, New Hampshire, USA

[2] Department of Pathology and Laboratory Medicine, Brown University, Providence, Rhode Island, USA

[3] Department of Public Health, Oregon State University, Corvallis, Oregon, USA

[4] Department of Pharmacology and Toxicology, Geisel School of Medicine at Dartmouth College, Hanover, New Hampshire, USA

Corresponding Author:

Carmen J. Marsit, PhD

Department of Pharmacology and Toxicology

Department of Community and Family Medicine

Geisel School of Medicine at Dartmouth College

7650 Remsen

Hanover, NH 03755

Phone: 603-650-1825

Fax: 603-650-1129

Carmen.J.Marsit@Dartmouth.edu

Conflict of Interest: E. Andres Houseman and Carmen J. Marsit, authors of this manuscript, are inventors on a pending international patent application, International Publication Number, WO 2012/162660, entitled "Method Using DNA Methylation for Identifying a Cell or a Mixture of Cells for Prognosis and Diagnosis of Diseases, and for Cell Remediation Therapies".

Abbreviations

CpG Island: A region with at least 200 bp, a GC percentage that is greater than 50%, and an observed-to-expected CpG ratio that is greater than 60%.

CpG Island Shore region: 0-2 kb from a CpG Island

CpG Island Shelf region: 2-4 kb from a CpG Island

Ocean region: > 4 kb from a CpG Island

DMA: Dimethylarsinic acid

DMR: Differentially methylation region

ER: Estrogen receptor

iAs: Inorganic arsenic

MMA: Monomethylarsonic acid

PBL: Peripheral blood leukocytes

U-As: Urinary arsenic

Abstract

Background: There is increasing epidemiologic evidence that arsenic exposure *in utero*, even at low levels found throughout much of the world, is associated with adverse reproductive outcomes and may contribute to long-term health effects. Animal models, *in vitro* studies, and human cancer data suggest that arsenic may induce epigenetic alterations, specifically by altering patterns of DNA methylation.

Objectives: This study aims to identify differences in DNA methylation in cord blood samples of infants with *in utero*, low-level arsenic exposure.

Methods: DNA methylation of cord-blood derived DNA from 134 infants involved in a prospective birth cohort in New Hampshire was profiled using the Illumina Infinium Methylation450K array. *In utero* arsenic exposure was estimated using maternal urine samples collected at 24-28 weeks gestation. We used a novel cell mixture deconvolution methodology for examining the association between inferred white blood cell mixtures in infant cord-blood and *in utero* arsenic exposure and also examined the association between methylation at individual CpG loci and arsenic exposure levels.

Results: We found an association between urinary inorganic arsenic concentration and the estimated proportion of CD8+ T-lymphocytes (1.18; 95% CI: 0.12, 2.23). Among the top 100 CpG loci with the lowest p-values based on their association with urinary arsenic levels, there was a statistically significant enrichment of these loci in CpG Islands ($p = 0.009$). Of those in CpG Islands ($n = 44$), the majority (75%) exhibited higher methylation levels in the highest exposed group compared to the lowest exposed group. There were also several CpG loci that exhibited linear dose-dependent relationship between methylation and arsenic exposure.

Conclusions: Our findings suggest that *in utero* exposure to low levels of arsenic may affect the epigenome. Long-term follow-up is planned to determine whether the observed changes are associated with health outcomes.

Introduction

Arsenic, an established environmental toxicant, has been linked to numerous adverse health outcomes, including cardiovascular disease (Chen et al. 2011; Gong and O'Bryant 2012), cancer (Hopenhayn-Rich et al. 1998), and skin lesions (Yunus et al. 2011). There is also a growing concern regarding the effects of *in utero* exposure to arsenic on the developing fetus (Vahter 2008). Recent studies have reported that high-level exposure to arsenic *in utero* is associated with increased infant mortality, low birth weight, and birth defects (Rahman et al. 2009; Rahman et al. 2010). Most previous studies of arsenic exposure and newborn health outcomes, however, have primarily focused on populations in arsenic endemic regions, outside of the United States (Vahter 2008). Whereas occupational exposure to arsenic and contaminated drinking water were once considered the primary means of arsenic exposure, dietary intake of arsenic, in particular from rice consumption, potentially may contribute to the same extent as drinking water in exposing pregnant women to arsenic (Gilbert-Diamond et al. 2011). This finding has raised concerns regarding the possible health consequences associated with fetal exposure to what may be relatively common levels of arsenic. Yet, there exists a gap in current knowledge as to whether findings from populations in arsenic endemic regions of the world can be used to infer the risks associated with low-level arsenic exposure, common to much of the United States.

Emerging data now suggest that arsenic may induce epigenetic alterations, specifically, by altering patterns of DNA methylation (Reichard and Puga 2010; Ren et al. 2011). Differences in global DNA methylation associated with exposure to arsenic have been reported based on animal (Davis et al. 2000; Zhao et al. 1997), *in vitro* (Mass and Wang 1997), and human studies across various tissue types (i.e., umbilical cord-blood and peripheral blood in adult subjects) (Intarasunanont et al. 2012; Wilhelm et al. 2010). The patterns of DNA methylation alteration

associated with arsenic exposure have been far from consistent, with several studies reporting reductions in global DNA methylation (Intarasunanont et al. 2012; Wilhelm et al. 2010; Zhao et al. 1997) and others reporting evidence of global hypermethylation (Davis et al. 2000; Mass and Wang 1997). Factors including the type of markers used for assessment of global methylation, the biological sample in which DNA methylation was measured, and modifiers such as nutritional folate levels (Pilsner et al. 2007) may explain the differential findings. In the context of prenatal exposure to arsenic, two recently published studies of different Bangladesh cohorts have reported positive associations between LINE-1 repeat element methylation in cord blood and arsenic exposure *in utero* (Kile et al. 2012; Pilsner et al. 2012), while no significant associations were observed in a cross-sectional study conducted in Thailand (Intarasunanont et al. 2012).

Beyond global methylation alterations, several recent epidemiologic studies have reported positive gene-specific associations between *in utero* exposure to arsenic and promoter methylation of *CDKN2A* (encoding p16INK4A) (Kile et al. 2012) and *TP53* (Intarasunanont et al. 2012) measured in human umbilical cord peripheral blood lymphocytes (PBLs). However, the reported associations of these two studies correspond to very modest differences in DNA methylation, even though these studies were comprised of individuals residing in the areas of the world with the highest purported levels of arsenic exposure. These studies as well as studies of adults, have measured DNA methylation in unfractioned PBLs, which are relatively easy to collect and process. However, the primary limitation is that methylation signatures in PBLs represent the aggregate methylation profile of a complex cellular mixture. Thus, even small changes in percent methylation may be indicative of considerable differences in underlying cell populations, reflecting immunomodulation. Arsenic exposure has been associated with immune

6

suppression and impaired macrophage function in exposed populations (Banerjee et al. 2009; Selgrade 2007). Hence, one possible mechanism for arsenic-induced diseases in exposed populations is altered immune function related to alterations of immune cell populations, which may be evident in blood-based profiles of DNA methylation. Dissecting out the contributions of different cell types and direct changes to the methylome on the observed associations between arsenic exposure and gene-specific DNA methylation patterns is critical for understanding the mechanisms of arsenic's immunotoxic effects. In order to address this limitation, work from our group has begun to use DNA methylation signatures of known cell types as a surrogate for defining cell mixture proportions (Houseman et al. 2012), and by applying this methodology we are able to interrogate not only arsenic's effect on DNA methylation but also its effect on relative leukocyte subtype proportions.

Despite the collective evidence supporting the role of arsenic exposure on the dysregulation of DNA methylation, little is known about potential effects of low-levels of arsenic *in utero* – common to much of the world's population - on DNA methylation. Moreover, the extent to which such patterns reflect immunomodulation, indicated by shifts in leukocyte subpopulations, or are representative of changes in the underlying methylome, has not been previously examined. We aimed to address these questions by examining the association between low-level arsenic exposure at 24-28 weeks gestation, a period of exposure during which DNA methylation patterns are becoming set in hematopoetic stem cells (Rodak 2007), and patterns of epigenome-wide DNA methylation in umbilical cord-blood samples from 134 mother/infant pairs enrolled in a US- based cohort study.

Methods

Study Population

The study population consisted of the 134 initial participants of the ongoing New Hampshire Birth Cohort Study (NHBCS), which is focused on pregnant women from New Hampshire, whose primary household drinking water source was a private well (Gilbert-Diamond et al. 2011). Eligibility criteria included English speaking, English literate, and mentally competent, pregnant women between 18-45 years old. Subjects who changed their residence since their last menstrual period or whose home water supply was from a source other than from a private well, were excluded from the study. Demographic and lifestyle information was collected during routine prenatal visits, and for the infant from the newborn medical chart. This study was approved by the Committee for the Protection of Human Subjects at Dartmouth College. All study participants provided written informed consent prior to the study.

Arsenic Exposure Assessment

As previously described (Gilbert-Diamond et al. 2011), spot urine samples were collected at approximately 24-28 weeks of gestation in acid washed containers that contained 30 μL of 10 mM diammonium diethyldithiocarbamate to stabilize arsenic species, and frozen at −80 °C until analysis (within 24 hours of collection). Samples were analyzed for individual species of urinary arsenic using a high-performance liquid chromatography (HPLC) ICP-MS system, and urinary creatinine levels were assessed to control for urinary dilution. The arsenic speciation method is capable of quantitatively determining five arsenic species in urine: As^{III}, As^{V}, DMA^{V}, MMA^{V}, and arsenobetaine. The separated arsenic species were detected by ICP-MS using time-resolved analysis at *m/z* 75. The detection limits ranged from 0.10 to 0.15 μg/L for the individual arsenic

species. Values for samples with measurements below the limit of detection were taken to be the median between 0 μg/L and the detection limit for that arsenic species. We calculated total urinary arsenic concentrations, hereafter referred to as U-As, by summing inorganic arsenic (As^{III} and As^{V}) and the metabolic products, MMA^{V} and DMA^{V}. Arsenobetaine was excluded from this calculation, as it is thought to be nontoxic and pass through the body without being metabolized. We utilized total U-As as a measure of *in utero* exposure to arsenic, as urinary arsenic levels have been suggested to provide reliable indications of internal dose (Marchiset-Ferlay et al. 2012) and arsenic is known to readily cross the placenta leading to fetal serum concentrations similar to maternal levels (Concha et al. 1998). As a measure of methylation efficiency, we have also calculated the ratio of inorganic to total urinary arsenic (iAs / (iAs + MMA^{V} + DMA^{V})).

DNA methylation assessment and quality control

DNA was isolated from cord-blood samples using DNeasy® blood & tissue kits (Qiagen, Valencia, CA) and bisulfite converted using the EZ DNA Methylation kit (Zymo, Irvine, CA). Samples were randomized across several plates and subsequently subjected to epigenome-wide DNA methylation assessment using the Illumina Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA), which simultaneously profiles the methylation status for > 485,000 CpG sites at single-nucleotide resolution. Microarrays were processed at the Biomedical Genomics Center at the University of Minnesota, following standard protocols. The methylation status for each individual CpG locus was calculated as the ratio of fluorescent signals (β = Max(M,0)/[Max(M,0)+Max(U,0) + 100]), ranging from 0 (no methylation) to 1 (complete methylation), using the average probe intensity for the methylated (M) and unmethylated (U) alleles. The data were assembled using BeadStudio methylation software (Illumina, San Diego,

9

CA), without normalization per the manufacturer's instructions. We used array control probes to assess the quality of our samples and evaluate potential problems such as poor bisulfite conversion or color-specific issues for each array (Marsit et al. 2009). All CpG loci on X and Y chromosomes and all loci within 100bp of known SNPs (determined using the annotation for the Illumina HumanMethylation450 array) were excluded from the analysis to avoid gender-specific methylation bias and biases related to genetic variability, respectively, leaving 385,249 autosomal CpG loci for analysis in 134 samples. Technical validation of the methylation array measurements was obtained using bisulfite pyrosequencing (see Supplemental Material, Bisulfite Pyrosequencing, for details).

Statistical Analysis

Principal components analysis and adjustment for plate-effects

DNA methylation values were logit (i.e., $\log(\beta/1-\beta)$) transformed as in previous studies (Du et al. 2010; Kuan et al. 2010). To de-convolve the most prevalent sources of variability in DNA methylation across the array, we performed a principal components analysis (PCA) on the resulting methylation data (Harper et al. 2013; Yang et al. 2010). PCA represents a feature extraction technique where the methylation data is orthogonally transformed, such that the first principal component has the largest possible variance (accounts for maximal amount of variability in the methylation data), and each succeeding component in turn has the next highest variance possible. The resulting top 3 principal components (those representing the maximum proportion of variability in methylation) were then examined in terms of their association with technical aspects concerning the array (i.e., plate/BeadChip) and patient demographic information using a series of linear regression models (see Supplemental Material, Figures S1-S2 and Table S1). As the top 3 principal components were significantly associated with plate

(Supplemental Material, Figures S2 and Table S1), suggesting that plate was a major source of variability in methylation across the array, we adjusted for plate-effects by applying the ComBat method (Johnson et al. 2007). Following a similar procedure, the resulting plate-adjusted methylation data were then investigated to ensure that variation in methylation across the array induced by plate-effects had been successfully attenuated (see Supplemental Material, Figure S2 and Table S2). The top 3 principal components computed from the plate-adjusted methylation data were also investigated in terms of their association with arsenic exposure. Briefly, this was accomplished by fitting a series of linear regression models that modeled the principal component as the dependent variable, quartiles of U-As as an independent variable, and were adjusted for maternal age at delivery, infant gender, and urinary creatinine levels (Barr et al. 2005; Gamble and Liu 2005). Unless stated otherwise, quartiles of U-As were used as measures of arsenic exposure as several studies have reported non-linear relationships between arsenic exposure and patterns of DNA methylation at both gene-specific and genome-wide levels (Chanda et al. 2006; Majumdar et al. 2010) and in order to reduce any skewing of the data by outlying values.

Cell mixture deconvolution analysis

Using the plate-adjusted methylation data, we employed a novel statistical methodology (Houseman et al. 2012) for inferring changes in the distribution of leukocytes between quartiles of U-As using DNA methylation signatures, in combination with a previously obtained external reference data set consisting of methylation signatures from purified leukocyte samples (Houseman et al. 2012; Koestler et al. 2012). Further details regarding the reference data set are provided elsewhere (Houseman et al. 2012; Koestler et al. 2012). A critical component of this approach is the set of DNA methylation signatures of the major leukocyte components of whole

blood (i.e., B cells, Natural Killer (NK) cells, CD8+ T Lymphocytes, CD4+T Lymphocytes, Monocytes, and Granulocytes). The methods of Houseman et al. 2012 demonstrate that the distribution of white blood cells can be approximated from the DNA methylation measured in whole blood at the top 500 leukocyte differentially methylated regions (L-DMRs); application of these methods allowed us to estimate the expected difference in cell type proportions between U-As quartiles 2, 3, and 4 and the referent quartile (quartile 1), as well as the expected change in cell type proportions based on a 1-μg/L increase in the concentration of the individual arsenic metabolites. In addition, the methods of Houseman et al. 2012 allowed us to quantify the proportion of total variability in cord-blood DNA methylation explained by estimated immune cell composition. Consistent with our other models, we adjusted for maternal age at delivery, infant gender and urinary creatinine levels. We note several assumptions regarding the cell mixture deconvolution method. First, we assume that the 500 L-DMRs, discerned using the methylation signature from purified leukocyte subtypes from healthy adults, are indeed biologically determinant of key differences in cell type across all ages, and therefore translate to cord-blood. We also assume that the methylation status for the 500 L-DMRs used here is not systematically altered by exposure to arsenic. We note other work (Koestler et al. under review) that demonstrates the insensitivity of the methods to the deletion of some CpGs (e.g. those that show evidence of serving as age-DMRs), i.e. that the accidental inclusion of some age-DMRs in the set are unlikely to influence results. We also note that substantial biases due to arsenic exposure would require systematic alterations of methylation states at the 500 L-DMRs in a manner consistent with the linear space spanned by the reference profiles, an event unlikely unless the alteration was itself implicated in perturbations of hematopoiesis.

Locus-locus analysis for detecting differentially methylated CpG loci

We implemented a locus-by-locus analysis aimed toward identifying differentially methylated CpG sites based on total urinary arsenic levels. Briefly, ANCOVA models were fit to each CpG site separately, and modeled logit-transformed methylation as the response against quartiles of total urinary arsenic (quartile 1 treated as the referent group). Models for U-As were adjusted for maternal age at delivery, infant gender, and urinary creatinine levels. Although our examination was exploratory in nature, p-values were adjusted for multiple comparisons by computing the Benjamini-Hochberg adjusted p-values (Benjamini and Hochberg 1995). CpG loci that exhibited a linear dose-response relationship were identified by fitting a series of linear regression models, which modeled U-As quartiles as a continuous covariate; U-As quartiles were assigned values of 1-4, for quartiles 1-4, respectively.

All analyses were carried out using the R statistical package, version 2.13 (www.r-project.org/).

Results

Demographic and clinical information for the 134 mother/infant pairs is provided in Table 1. Home tap water arsenic concentrations ranged from close to the detection limit (0.03 µg/L) to nearly 100 µg/L, with twenty-two women (18%) having home drinking water above 10 µg/L, the current US Environmental Protection Agency (US EPA) standard. Total U-As, calculated by summing the concentrations of inorganic arsenic, monomethylarsonic acid (MMA), and dimethylarsinic acid (DMA), ranged from (0.45 µg/L) to nearly 300 µg/L, was highly right-skewed (see Supplemental Material, Figure S3), and had a median concentration of 4.1 µg/L (IQR 1.8 − 6.6), which is similar to the median estimated for the US population as a whole (Caldwell et al. 2009). A single outlying sample with a urinary As value of nearly 300 µg/L was

confirmed. Due to the skew of the data and this outlier, as well as the potential non-linear relationship between arsenic exposure and DNA methylation, we examined total urinary arsenic in all analyses as a categorical variable splitting the data in quartiles.

Supplemental Material, Table S3 presents coefficient estimates and corresponding p-values for quartiles of total U-As and their association with shifts in estimated lymphocyte subsets. Coefficient estimates for each cell type reflect the estimated percent difference in the proportion of that cell type between U-As quartiles 2-4 relative to the lowest quartile (Quartile 1). There were no statistically significant overall differences in the proportions of cell types among quartiles of total U-As. Table 2 provides the results of this model fit to continuous values of total urinary inorganic As and metabolites individually where the coefficient estimates represent the estimated change in the proportion of a particular cell type based on a 1-μg/L increase in the levels of these metabolites. For iAs / (iAs + MMA^V + DMA^V), the coefficient estimates represent the expected difference in cell type proportions for a 1-unit increase in the proportion of inorganic As to total U-As. There was a statistically significant positive association between a 1-μg/L increase in inorganic arsenic and the proportion of CD8+ T-lymphocytes (1.18; 95% CI: 0.12, 2.23) (Table 2). CD8+ T-lymphocytes also were associated with a 1-unit increase in the ratio of inorganic arsenic to total urinary arsenic (9.11; 95% CI: 0.44, 17.79), but were not significantly associated with other fractions (i.e., MMA^V and DMA^V, Table 2).

Overall however, WBC distributions explained a relatively small proportion of the variability in patterns of cord-blood DNA methylation associated with *in utero* exposure to arsenic (3% for total U-As), and so the remaining systematic source of variation was hypothesized to be alterations to the methylome itself in some or all of the cell populations examined. We first used a principal components analysis (PCA) to reduce the dimensionality of the genome-wide DNA

14

methylation data and to investigate the extent to which low-level arsenic exposure *in utero* was associated with genome-wide variability in cord-blood DNA methylation in our population. The results of this analysis (see Supplemental Material, Table S4) revealed no statistically significant association between U-As quartiles and principal components 1-3, which were estimated to account for 20%, 12% and 9% of the variation in methylation across the array.

To identify individual CpG loci whose methylation status was associated with exposure, we next examined differential locus-specific patterns of DNA methylation based on total U-As levels. As shown in Figure 1A, which depicts $-\log_{10}$(p-values) (y-axis) for the association between U-As quartiles and the methylation of individual CpG loci (the dependent variable), 68,353 of 385,249 CpG loci (18%) were significantly associated with exposure ($p < 0.05$ for a difference over all quartiles of U-As), controlling for maternal age at delivery, infant gender, and urinary creatinine levels. However no association was statistically significant after adjusting for multiple comparisons. Among the 100 CpG loci with the smallest unadjusted p-values for the difference in methylation over all U-As quartiles (see Supplemental Material, Table S5), there was a statistically significantly greater than expected proportion of loci located in CpG Islands (Fisher's exact $p = 0.009$) (Figure 1B), and of those in CpG Islands ($n = 44$), the majority (75%) exhibited higher methylation levels in the highest exposed group (quartile 4) compared to the lowest exposed group (quartile 1). Furthermore, a number of differentially methylated CpG loci displayed a linear dose-response relationship across the quartiles of total U-As levels (see Supplemental Material, Table S6). Most notably, our analysis identified *cg08884395* and *cg27514608,* associated with genes *ESR1* (Estrogen receptor 1) and *PPARGC1A* (Peroxisome proliferator-activated receptor gamma coactivator 1-alpha), respectively (Figure 2), which both exhibited a negative association between methylation and arsenic exposure (linear trend $p =$

0.0009 for both). Considering the differences in toxicity of inorganic arsenic, MMA and DMA (Le et al. 2000), exposure to different forms of arsenic over the range of U-As concentrations could potentially confound our results. To examine the robustness of our results based on exposure to different forms of arsenic, we refit models to the CpG sites that exhibited a dose-response relationship and individually adjusted for urinary concentrations of inorganic arsenic, MMA, and DMA. These results (see Supplemental Material, Tables S7-S9) indicated that adjustment for urinary arsenic metabolites had very little effect on the association between quartiles of U-As and the DNA methylation of the dose-response CpG sites, with all p-values for trend remaining less than 0.05.

As validation of the methylation array, two CpG loci (*cg27308738*, n=29 and *cg10528424*, n=30) were subjected to bisulfite pyrosequencing, as these loci were among the most variable across the study samples in terms of their methylation levels. Our results indicated a very high degree of correlation between the methylation array values and the percent methylation measurements obtained from pyrosequencing (Pearson's correlation coefficient = 0.98 and 0.93 for *cg27308738* and *cg10528424*, respectively) (see Supplemental Material, Figure S4).

Discussion

Patterns of DNA methylation are established during embryogenesis and play an important role in gene transcription, chromosomal stability, X-chromosome inactivation, and tissue differentiation. Alteration of fetal DNA methylation is a potential mechanism linking *in utero* exposures to chronic diseases in adulthood (Kile et al. 2012). The predominantly rural US study population is distinct from prior study populations, which were mainly based outside of the US and were typically exposed to greater levels of arsenic than the study population examined here. However,

nearly 20% of this pregnant population was exposed to household drinking water above the US EPA standard. Not surprisingly, the majority of women with household drinking water above the US EPA standard were among those with the highest urinary arsenic levels (data not shown). This is in line with our previous report (Gilbert-Diamond et al. 2011) that arsenic exposure from private well water contributes appreciably to exposure in this population. Our results, which indicated some evidence of differential patterns of DNA methylation across the quartiles of arsenic exposure, support additional investigation into the biologic effects of this level of arsenic exposure and provide further evidence for the need of public health efforts to reduce these exposures.

Animal and *in vitro* studies have demonstrated effects of arsenic at environmentally relevant levels (i.e., 10 -100 µg/L) on early immune response (Kozul et al. 2009a; Kozul et al. 2009b; Martin-Chouly et al. 2011; Mattingly et al. 2009), providing a potential biological mechanism for health complications experienced by prenatally exposed arsenic individuals later in life. Here, we examined whether low-level arsenic exposure between 24-28 weeks gestation, a period during which fetal hematopoiesis is shifting from the liver to the bone marrow (Rodak 2007), was associated with the shifts in immune cells. Our results revealed a statistically significant positive association between the estimated proportion of CD8+ T-lymphocytes and both inorganic arsenic and the proportion of inorganic arsenic to total urinary arsenic. Although the methylated species of arsenic also exhibited a positive association with CD8+ T-lymphocytes, unlike inorganic arsenic, the association between MMA and DMA and CD8+ T-lymphocytes did not meet statistical significance (Table 2). Concentrations of individual metabolites reflect exposure to inorganic arsenic, direct consumption of methylated forms, and methylation efficiency (Kile et al. 2009). Our findings of a stronger association with inorganic arsenic may

be in part due to the varying toxicity of arsenic's metabolites or epigenetic activity of arsenic's metabolites. Thus, our results indicate that increased levels of urinary inorganic arsenic, (due to greater *in utero* exposure and/or potentially reduced metabolism of inorganic arsenic to the methylated species), are associated with increased proportions of cord-blood CD8+ T-lymphocytes. These results are consistent with recent work in which adult mice exposed to 100 μg/L of arsenic over a five-week period showed increased percentage and total levels of CD8+ T-lymphocytes (Kozul et al. 2009a). In humans though, a recent report that evaluated the effects of prenatal arsenic exposure on thymus function at birth in a Bangledeshi cohort of 130 mother/infant pairs raised the possibility that there may be functional deficits in CD8+ cells associated with arsenic exposure *in utero* (Ahmed et al. 2012). Although we recognize our findings are preliminary, we know of no prior studies reporting on relationships between arsenic exposure and cord-blood immune cell proportions defined using DNA methylation profiles as a surrogate measure of leukocyte proportion. The potential importance of these findings is underscored by mounting scientific evidence demonstrating that CD8+ T-lymphocytes contribute to the initiation, progression, and regulation of several pathogenic autoimmune responses (Walter and Santamaria 2005), possibly providing a mechanism by which *in utero* exposure to arsenic during relevant etiologic periods of hematopoiesis results in an increased risk of autoimmune related illness later in life (Tseng 2004). It will be critical to examine prospectively if the alterations we have observed are linked to immunological consequences later in childhood as well as to validate our findings in additional cohorts.

We further investigated the extent to which low-level arsenic exposure *in utero* was associated with genome-wide DNA methylation variability, assessed using principal components based on DNA methylation in cord-blood. This analysis revealed no significant associations between the

top three principal components and quartiles of U-As, suggesting that arsenic exposure may not be leading to overt, genome-wide perturbations in DNA methylation extent which had been suggested by studies of "global methylation" markers. Instead, we found examining associations between arsenic and each CpG loci individually, that 18% of the CpG loci tested were differentially methylated among quartiles of U-As (p <0.05). However none of these associations remained significant after adjustment for multiple comparisons. We observed that the top 100 CpGs associated with total urinary arsenic on the basis of unadjusted p-values were disproportionately located in CpG islands, and the majority of these CpG loci had higher methylation levels in the highest exposed group (quartile 4) compared to the lowest exposed group (quartile 1). Such findings are in line with existing reports, which assert that arsenic exposure is associated with hypermethylation of promoter region CpG islands (Chen et al. 2001) and the well-established role of promoter CpG Island methylation in epigenetic gene control and disease states (Ferreira et al. 2012; Lorenzen et al. 2012). Additionally, there were a number of CpG sites that exhibited a linear dose-response relationship with respect to quartiles of U-As (see Supplemental Material, Table S6), and those relationships remained robust to individual adjustment for the levels of arsenic's metabolites (see Supplemental Material, Table S7-S9). While these results require further validation, they suggest that even low-level *in utero* exposure to arsenic may bring about gene or CpG specific epigenetic changes, which themselves may contribute to altered gene expression and downstream cellular function.

In addition to purported effects of arsenic as a carcinogen and immunotoxicant, a growing body of literature is demonstrating endocrine disruption as an additional mode of toxicity (Naujokas et al. 2013). We noted in our analyses that increased levels of total urinary arsenic was associated with decreased methylation of *cg08884395* (linear trend p = 0.0009) (Figure 2A, Supplemental

Table S6), located in CpG island shore region of *ESR1*, encoding estrogen receptor α. Importantly, methylation of this gene was observed in a CpG island shore region - regions up to 2 kb flanking CpG Island - which have been suggested to be the most enriched with functional CpG sites (Irizarry et al. 2009). In addition, although not statistically significant after adjustment for multiple comparisons, our analysis revealed a decrease in the methylation of *cg27514608*, for increasing quartiles of U-As (linear trend $p = 0.0009$) (Figure 2B, Supplemental Table S6). This locus is associated with the gene *PPARGC1A*, encoding PGC-1α, a coactivator of several nuclear receptors including peroxisome proliferator-activated receptors α and γ, thyroid hormone receptor, mineral corticoid receptor, and ERs (Sugawara et al. 2001). The result with *ESR1* is consistent with a report of decreased methylation and increased expression of *ESR1* with *in vitro* arsenic exposure in a breast cancer model (Du et al. 2012), while the association with *PPARGC1A* is consistent with a report demonstrating upregulation of this gene in arsenic-induced Bowen's disease, a skin carcinoma (Lee et al. 2011). Given the role of hormone receptors and their pathways in fetal development, and throughout childhood, and the role that arsenic may be playing in altering these pathways, our data suggest that epigenetic modes of altered regulation of endocrine pathways may also be important for future study.

There are notable limitations to the present study. Our study cohort is largely Caucasian, which could limit generalizability. However, our focus on common exposure levels is unique compared to studies based in arsenic endemic regions. While our study did not directly assess concentrations of arsenic on umbilical cord-blood as a measure of fetal exposure to arsenic, arsenic is known to readily cross the placenta leading to fetal serum concentrations similar to maternal levels (Concha et al. 1998; Hall et al. 2007). We also note that the external reference set used in our cell mixture analysis, which consisted of DNA methylation signatures from

purified WBCs, was isolated from different, anonymous, adult, non-diseased individuals' whole blood (Koestler et al. 2012). As differences in DNA methylation have been demonstrated in individuals of varying ages (Winnefeld and Lyko 2012), the WBC-specific methylation signatures may not accurately reflect cell-specific patterns of DNA methylation evident in cord-blood, which is primarily comprised of infant blood. We further note that our cell mixture analysis used signatures of DNA methylation as a surrogate for cell mixture composition, while a complete assessment of the immune profile would require flow cytometric measurements. While the results from our locus-by-locus analysis showed some evidence of altered methylation with arsenic exposure, these results did not withstand multiple comparisons corrections. Thus, these results need to be replicated in additional studies. Along the same lines, we do not have the power to consider differential effects by sex or other potential modifiers, but larger studies should consider such effects.

Conclusions

Our novel exploratory examination suggests that low-level arsenic exposure *in utero* may influence the infant epigenome. Particularly, our findings highlight the possibility that exposure to arsenic – even at levels common to much of the world's population - during relevant etiologic periods of fetal development may induce shifts in underlying cell populations, as well as gene specific alterations in DNA methylation. Long-term follow-up is planned to determine whether the observed changes are associated with short- and long-term health outcomes.

References

Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, et al. 2012. Heterogeneity in white blood cells has potential to confound DNA methylation measurements. PLoS One 7:e46705.

Ahmed S, Ahsan KB, Kippler M, Mily A, Wagatsuma Y, Hoque AM, et al. 2012. In utero arsenic exposure is associated with impaired thymic function in newborns possibly via oxidative stress and apoptosis. Toxicol Sci.

Banerjee N, Banerjee S, Sen R, Bandyopadhyay A, Sarma N, Majumder P, et al. 2009. Chronic arsenic exposure impairs macrophage functions in the exposed individuals. J Clin Immunol 29:582-594.

Barr DB, Wilder LC, Caudill SP, Gonzalez AJ, Needham LL, Pirkle JL. 2005. Urinary creatinine concentrations in the u.S. Population: Implications for urinary biologic monitoring measurements. Environ Health Perspect 113:192-200.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 57:289-300.

Caldwell KL, Jones RL, Verdon CP, Jarrett JM, Caudill SP, Osterloh JD. 2009. Levels of urinary total and speciated arsenic in the us population: National health and nutrition examination survey 2003-2004. J Expo Sci Environ Epidemiol 19:59-68.

Chanda S, Dasgupta UB, Guhamazumder D, Gupta M, Chaudhuri U, Lahiri S, et al. 2006. DNA hypermethylation of promoter of gene p53 and p16 in arsenic-exposed people with and without malignancy. Toxicol Sci 89:431-437.

Chen H, Liu J, Merrick BA, Waalkes MP. 2001. Genetic events associated with arsenic-induced malignant transformation: Applications of cdna microarray technology. Mol Carcinog 30:79-87.

Chen Y, Graziano JH, Parvez F, Liu M, Slavkovich V, Kalra T, et al. 2011. Arsenic exposure from drinking water and mortality from cardiovascular disease in bangladesh: Prospective cohort study. BMJ 342:d2431.

Concha G, Vogler G, Lezcano D, Nermell B, Vahter M. 1998. Exposure to inorganic arsenic metabolites during early human development. Toxicol Sci 44:185-190.

Davis CD, Uthus EO, Finley JW. 2000. Dietary selenium and arsenic affect DNA methylation in vitro in caco-2 cells and in vivo in rat liver and colon. J Nutr 130:2903-2909.

Du J, Zhou N, Liu H, Jiang F, Wang Y, Hu C, et al. 2012. Arsenic induces functional re-expression of estrogen receptor alpha by demethylation of DNA in estrogen receptor-negative human breast cancer. PLoS One 7:e35957.

Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. 2010. Comparison of beta-value and m-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 11:587.

Ferreira HJ, Heyn H, Moutinho C, Esteller M. 2012. Cpg island hypermethylation-associated silencing of small nucleolar rnas in human cancer. RNA Biol 9.

Gamble MV, Liu X. 2005. Urinary creatinine and arsenic metabolism. Environ Health Perspect 113:A442; author reply A442-443.

Gilbert-Diamond D, Cottingham KL, Gruber JF, Punshon T, Sayarath V, Gandolfi AJ, et al. 2011. Rice consumption contributes to arsenic exposure in us women. Proc Natl Acad Sci U S A 108:20656-20660.

Gong G, O'Bryant SE. 2012. Low-level arsenic exposure, as3mt gene polymorphism and cardiovascular diseases in rural texas counties. Environ Res 113:52-57.

Hall M, Gamble M, Slavkovich V, Liu X, Levy D, Cheng Z, et al. 2007. Determinants of arsenic metabolism: Blood arsenic metabolites, plasma folate, cobalamin, and homocysteine concentrations in maternal-newborn pairs. Environ Health Perspect 115:1503-1509.

Harper KN, Peters BA, Gamble MV. 2013. Batch effects and pathway analysis: Two potential perils in cancer studies involving DNA methylation array analysis. Cancer Epidemiol Biomarkers Prev.

Hopenhayn-Rich C, Biggs ML, Smith AH. 1998. Lung and kidney cancer mortality associated with arsenic in drinking water in cordoba, argentina. Int J Epidemiol 27:561-569.

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13:86.

Intarasunanont P, Navasumrit P, Woraprasit S, Chaisatra K, Suk WA, Mahidol C, et al. 2012. Effects of arsenic exposure on DNA methylation in cord blood samples from newborn babies and in a human lymphoblast cell line. Environ Health 11:31.

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific cpg island shores. Nat Genet 41:178-186.

Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics 8:118-127.

Kile ML, Hoffman E, Hsueh YM, Afroz S, Quamruzzaman Q, Rahman M, et al. 2009. Variability in biomarkers of arsenic exposure and metabolism in adults over time. Environ Health Perspect 117:455-460.

Kile ML, Baccarelli A, Hoffman E, Tarantini L, Quamruzzaman Q, Rahman M, et al. 2012. Prenatal arsenic exposure and DNA methylation in maternal and umbilical cord blood leukocytes. Environ Health Perspect.

Koestler DC, Marsit CJ, Christensen BC, Accomando WP, Langevin SM, Houseman EA, et al. 2012. Peripheral blood immune cell methylation profiles are associated with non-hematopoietic cancers. Cancer Epidemiol Biomarkers Prev.

Kozul CD, Ely KH, Enelow RI, Hamilton JW. 2009a. Low-dose arsenic compromises the immune response to influenza a infection in vivo. Environ Health Perspect 117:1441-1447.

Kozul CD, Hampton TH, Davey JC, Gosse JA, Nomikos AP, Eisenhauer PL, et al. 2009b. Chronic exposure to arsenic in the drinking water alters the expression of immune response genes in mouse lung. Environ Health Perspect 117:1108-1115.

Kuan PF, Wang S, Zhou X, Chu H. 2010. A statistical framework for illumina DNA methylation arrays. Bioinformatics 26:2849-2855.

Le XC, Ma M, Cullen WR, Aposhian HV, Lu X, Zheng B. 2000. Determination of monomethylarsonous acid, a key arsenic methylation intermediate, in human urine. Environ Health Perspect 108:1015-1018.

Lee CH, Wu SB, Hong CH, Liao WT, Wu CY, Chen GS, et al. 2011. Aberrant cell proliferation by enhanced mitochondrial biogenesis via mttfa in arsenical skin cancers. Am J Pathol 178:2066-2076.

Lorenzen JM, Martino F, Thum T. 2012. Epigenetic modifications in cardiovascular disease. Basic Res Cardiol 107:245.

Majumdar S, Chanda S, Ganguli B, Mazumder DN, Lahiri S, Dasgupta UB. 2010. Arsenic exposure induces genomic hypermethylation. Environ Toxicol 25:315-318.

Marchiset-Ferlay N, Savanovitch C, Sauvant-Rochat MP. 2012. What is the best biomarker to assess arsenic exposure via drinking water? Environ Int 39:150-171.

Marsit CJ, Christensen BC, Houseman EA, Karagas MR, Wrensch MR, Yeh RF, et al. 2009. Epigenetic profiling reveals etiologically distinct patterns of DNA methylation in head and neck squamous cell carcinoma. Carcinogenesis 30:416-422.

Martin-Chouly C, Morzadec C, Bonvalet M, Galibert MD, Fardel O, Vernhet L. 2011. Inorganic arsenic alters expression of immune and stress response genes in activated primary human t lymphocytes. Mol Immunol 48:956-965.

Mass MJ, Wang L. 1997. Arsenic alters cytosine methylation patterns of the promoter of the tumor suppressor gene p53 in human lung cells: A model for a mechanism of carcinogenesis. Mutat Res 386:263-277.

Mattingly CJ, Hampton TH, Brothers KM, Griffin NE, Planchart A. 2009. Perturbation of defense pathways by low-dose arsenic exposure in zebrafish embryos. Environ Health Perspect 117:981-987.

Naujokas MF, Anderson B, Ahsan H, Aposhian HV, Graziano JH, Thompson C, et al. 2013. The broad scope of health effects from chronic arsenic exposure: Update on a worldwide public health problem. Environ Health Perspect 121:295-302.

Pilsner JR, Liu X, Ahsan H, Ilievski V, Slavkovich V, Levy D, et al. 2007. Genomic methylation of peripheral blood leukocyte DNA: Influences of arsenic and folate in bangladeshi adults. Am J Clin Nutr 86:1179-1186.

Pilsner JR, Hall MN, Liu X, Ilievski V, Slavkovich V, Levy D, et al. 2012. Influence of prenatal arsenic exposure and newborn sex on global methylation of cord blood DNA. PLoS One 7:e37147.

Rahman A, Vahter M, Smith AH, Nermell B, Yunus M, El Arifeen S, et al. 2009. Arsenic exposure during pregnancy and size at birth: A prospective cohort study in bangladesh. Am J Epidemiol 169:304-312.

Rahman A, Persson LA, Nermell B, El Arifeen S, Ekstrom EC, Smith AH, et al. 2010. Arsenic exposure and risk of spontaneous abortion, stillbirth, and infant mortality. Epidemiology 21:797-804.

Reichard JF, Puga A. 2010. Effects of arsenic exposure on DNA methylation and epigenetic gene regulation. Epigenomics 2:87-104.

Ren X, McHale CM, Skibola CF, Smith AH, Smith MT, Zhang L. 2011. An emerging role for epigenetic dysregulation in arsenic toxicity and carcinogenesis. Environ Health Perspect 119:11-19.

Rodak BFF, G.A.; Doig, K 2007. Hematology: Clinical principles and applications. 3 ed:Saunder.

Selgrade MK. 2007. Immunotoxicity: The risk is real. Toxicol Sci 100:328-332.

Sugawara A, Takeuchi K, Uruno A, Ikeda Y, Arima S, Kudo M, et al. 2001. Transcriptional suppression of type 1 angiotensin ii receptor gene expression by peroxisome proliferator-activated receptor-gamma in vascular smooth muscle cells. Endocrinology 142:3125-3134.

Tseng CH. 2004. The potential biological mechanisms of arsenic-induced diabetes mellitus. Toxicol Appl Pharmacol 197:67-83.

Vahter M. 2008. Health effects of early life exposure to arsenic. Basic Clin Pharmacol Toxicol 102:204-211.

Walter U, Santamaria P. 2005. Cd8+ t cells in autoimmunity. Curr Opin Immunol 17:624-631.

Wilhelm CS, Kelsey KT, Butler R, Plaza S, Gagne L, Zens MS, et al. 2010. Implications of line1 methylation for bladder cancer risk in women. Clin Cancer Res 16:1682-1689.

Winnefeld M, Lyko F. 2012. The aging epigenome: DNA methylation from the cradle to the grave. Genome Biol 13:165.

Yang HH, Hu N, Wang C, Ding T, Dunn BK, Goldstein AM, et al. 2010. Influence of genetic background and tissue types on global DNA methylation patterns. PLoS One 5:e9355.

Yunus M, Sohel N, Hore SK, Rahman M. 2011. Arsenic exposure and adverse health effects: A review of recent findings from arsenic and health studies in matlab, bangladesh. Kaohsiung J Med Sci 27:371-376.

Zhao CQ, Young MR, Diwan BA, Coogan TP, Waalkes MP. 1997. Association of arsenic-induced malignant transformation with DNA hypomethylation and aberrant gene expression. Proc Natl Acad Sci U S A 94:10907-10912.

Table 1: Demographic, clinical, and phenotypic information for the study population.

Values are presented as mean (Adalsteinsson et al.), number (%), or median (interquartile range).

Characteristic	Value
Number of mother child pairs	134
Maternal age at delivery (yrs)	31 (4.6)
Gestational age (wks)	40 (1.3)
Birth weight (kg)	3.5 (0.46)
Infant gender	
Male	67 (50%)
Female	67 (50%)
Delivery type	
Vaginal	92 (69%)
C-Section	40 (30%)
Unknown	2 (1%)
Smoking Status	
Never	97 (72%)
Former	11 (8%)
Current	4 (3%)
Unknown	22 (16%)
School level	
Less than 11th grade	2 (1%)
High school graduate or equivalent	11 (8%)
Junior college graduate or some college or technical school	30 (22%)
College graduate	48 (36%)
Post-graduate schooling	21 (16%)
Unknown	22 (16%)
Maternal Urinary As (µg/L)	4.1 (1.8-6.6)
iAs (µg/L)[a]	0.25 (0.13-0.47)
MMA^V (µg/L)	0.31 (0.15-0.5)
DMA^V (µg/L)	3.4 (1.6-5.7)
Tap Water Arsenic (µg/L)	1.2 (0.2-6.2)

[a] 56 samples had measurements below the limit of detection for iAs.

Table 2: Estimated change in the proportions of leukocyte types by continuous specific urinary arsenic exposure measures. Results are based on fitting the cell mixture deconvolution method separately for each arsenic measure (i.e., iAs, MMA^V, ect.). Given below are the coefficient estimates and (95% CIs), where the coefficient estimates represent the estimated change in the proportion of a particular cell type based on a 1-μg/L increase in the levels of these metabolites. For iAs / (iAs + MMA^V + DMA^V), the coefficient estimates represent the

| | Estimated Change in Cell Type Proportion (95% CI) | | | | | |
| | Lymphocytes | | | | Myeloid cells | |
Arsenic Measure	CD8+-T	CD4+-T	NK cells	B-cells	Monocytes	Granulocytes
iAs (per μg/L)	1.18 (0.12, 2.23)*	-1.24 (-3.15, 0.68)	-0.11 (-1.83, 1.62)	-0.78 (-1.91, 0.36)	-0.37 (-1.31, 0.56)	1.38 (-2.30, 5.06)
MMA^V (per μg/L)	0.93 (-0.30, 2.15)	-0.24 (-2.62, 2.14)	-0.48 (-2.59, 1.62)	-0.68 (-1.88, 0.52)	-0.20 (-1.33, 0.93)	1.15 (0.53, -3.02, 5.32)
DMA^V (per μg/L)	0.42 (-0.80, 1.64)	-0.10 (-2.40, 2.20)	-0.37 (-2.14, 1.41)	-0.22 (-1.46, 1.01)	-0.27 (-1.16, 0.62)	0.40 (-3.57, 4.36)
iAs / (iAs + MMA^V + DMA^V)	9.11 (0.44, 17.79)*	-11.82 (-27.66, 4.02)	-2.16 (-14.58, 10.27)	-6.05 (-16.4, 4.27)	-1.81 (-8.20, 4.58)	16.91 (-14.12, 47.92)

Notes: All four models were controlled for maternal age at delivery, infant gender, and urinary creatinine. * Indicates p ≤ 0.05

Figure Legends

Figure 1: Locus-by-locus examination of differentially methylated CpG sites based on U-As levels. (A) Manhattan plot for total U-As, where points represent the $-\log_{10}$(p-value) testing the null hypothesis of no difference in methylation across quartiles of arsenic exposure, adjusted for maternal age at delivery, infant gender, and urinary creatinine levels. Red and blue lines indicate $-\log_{10}(1 \times 10^{-4})$ and $-\log_{10}(0.05)$, respectively. (B) Location of the top 100 CpGs associated with U-As on the basis of p-value (Top 100 probes) compared to all CpGs on the methylation array (All probes). N. Shore and N. Shelf refer to CpG Island Shore and Shelf regions, respectively that are upstream of a CpG Island region. S. Shore and S. Shelf refer to CpG Island Shore and Shelf regions, respectively, that are downstream of a CpG Island region

Figure 2: Crude plots of DNA methylation for (A) cg08884395 and (B) cg27514608 by quartiles of total U-As for the 134 study subjects. Red lines denote the within quartile median methylation status.
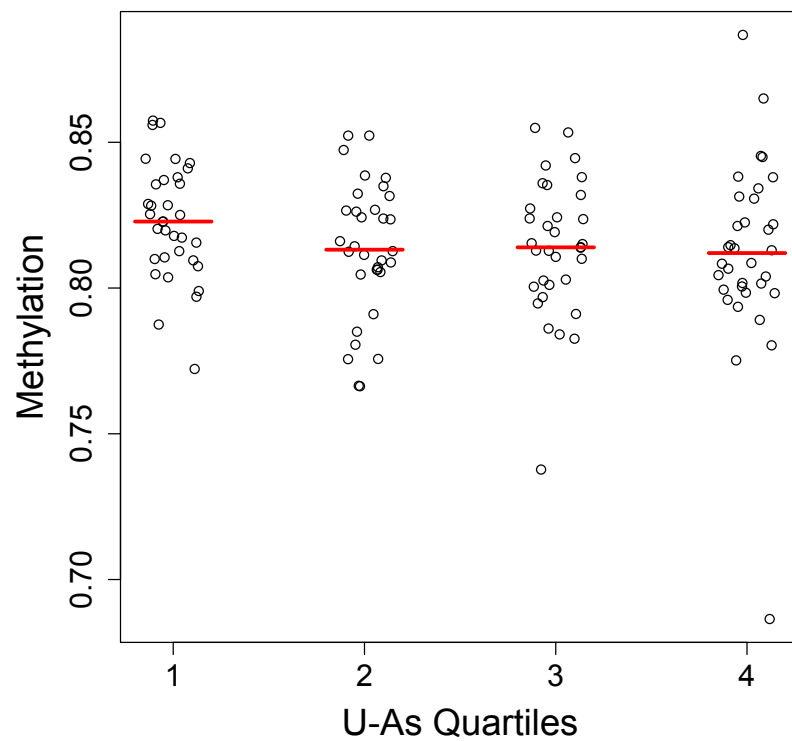
A

B
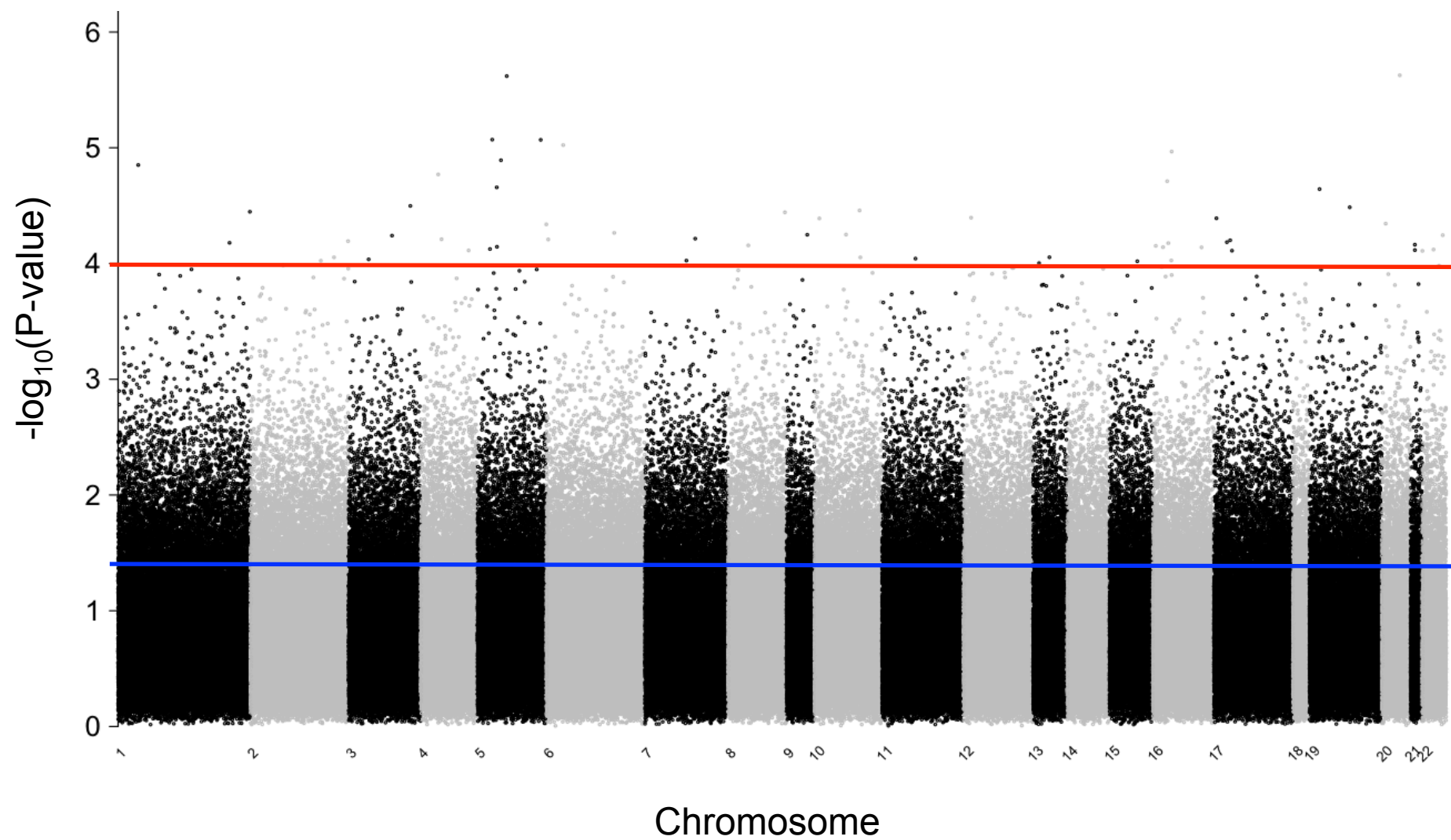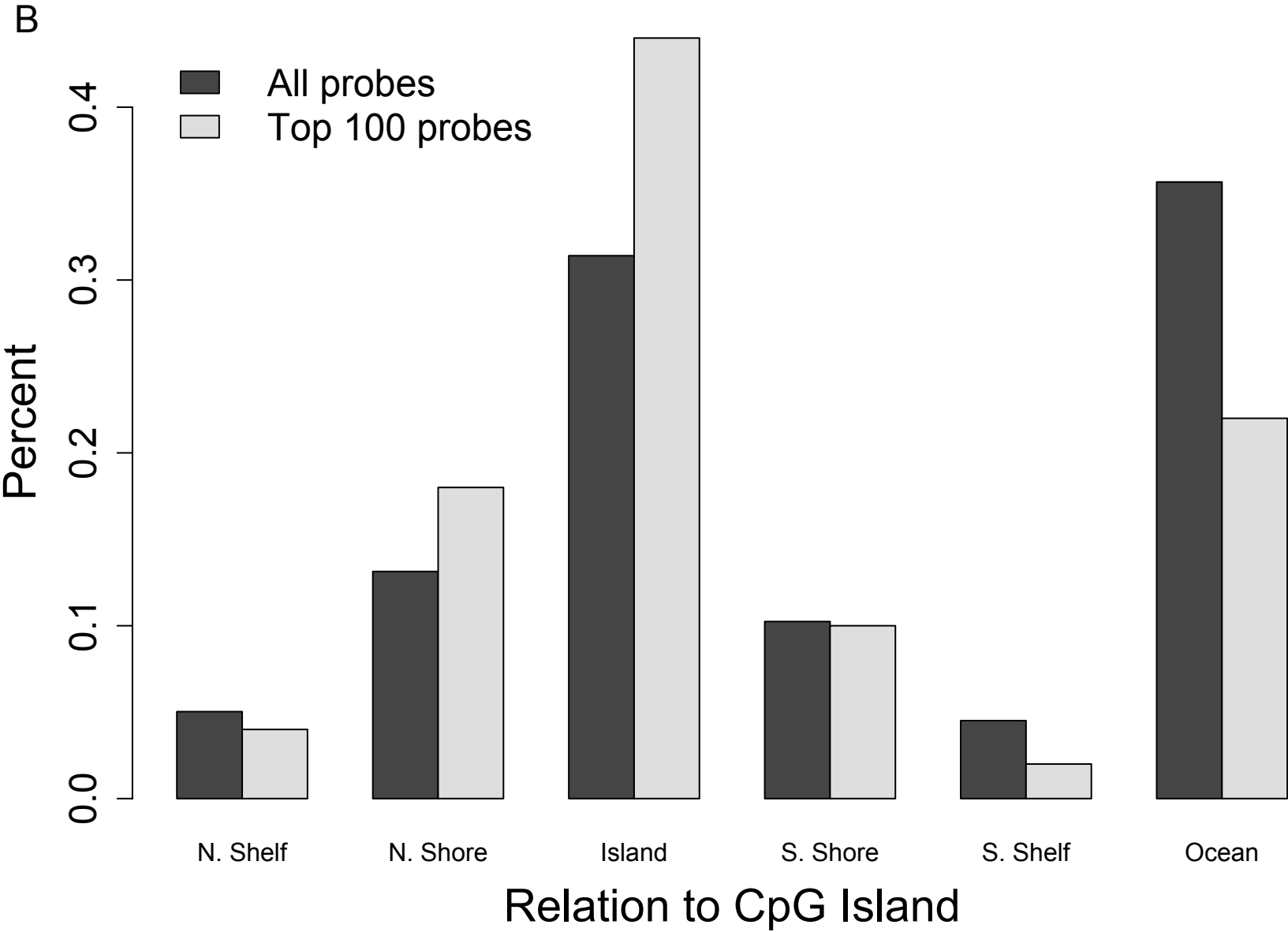
A   cg08884395 (ESR1)

B   cg27514608 (PPARGC1A)